

# Developing an Interdisciplinary Data Science Program

Mariam Salloum, Daniel Jeske, Wenxiu Ma, Vagelis Papalexakis, Christian Shelton, Vassilis Tsotras, and Shuheng Zhou\*

(mariam.salloum,daniel.jeske,wenxiu.ma,epapalex,cshelton,vtsotras,simone.zhou)@ucr.edu  
UC Riverside  
Riverside, CA, USA

## ABSTRACT

This paper describes a newly developed Data Science major that is jointly offered between two departments (Computer Science and Statistics) and across two colleges (College of Engineering and College of Natural and Architectural Sciences). The paper explores the inspiration and motivation for creating the undergraduate Data Science program, the program curriculum, and the governing body of the program. We aim to recruit undergraduates, specifically women, into the new major which are underrepresented in Computer Science at our University. We expect that this Major in Data Science will serve as a general framework for establishing similar interdisciplinary programs.

## CCS CONCEPTS

• **Social and professional topics** → **Model curricula**; **Computer science education**; **Model curricula**.

## KEYWORDS

Computational science education, Data Science, curriculum, interdisciplinary major, STEM

### ACM Reference Format:

Mariam Salloum, Daniel Jeske, Wenxiu Ma, Vagelis Papalexakis, Christian Shelton, Vassilis Tsotras, and Shuheng Zhou. 2021. Developing an Interdisciplinary Data Science Program. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21), March 13–20, 2021, Virtual Event, USA*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3408877.3432454>

## 1 INTRODUCTION

We live in a world where data is being generated continuously by scientific experiments, digital processes, sensors, social media, mobile devices, etc. The term ‘big data’ refers to data that is arriving from multiple sources at an alarming volume, velocity, and variety. There are increasing numbers of “small” datasets too. Training professionals to handle it properly, and not overfit or draw erroneous conclusions, is also important. Data Science deals with the extraction of knowledge from big data. As a scientific field,

Data Science affects research in many domains, including biological sciences, physical sciences, social sciences and humanities. The White House ‘Big Data Research and Development Initiative’ [3] committed \$200 million to ‘extract knowledge and insights from large and complex collections of digital data, accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning.’ NIH launched the Big Data To Knowledge (BD2K) [4] initiative ‘to enable biomedical research as a digital research enterprise, to facilitate discovery and support new knowledge.’ Other funding agencies (NSF, DARPA, IARPA, etc.) have similar research initiatives.

In addition to research, Data Science heavily influences economics and business. Data has become ubiquitous in everyday life; it impacts every profession, from entry-level office workers to CEOs, from team coaches to general managers, from accountants to CFOs. Businesses now have data available to at a scale that is historically unprecedented; harnessing this data for insight on what customers want provides them with a competitive advantage. This has led to a very high demand for Data Scientists. Such jobs require expertise in both computing and statistical methods.

Currently there is a plethora of M.S. programs (many of them on-line) in Data Science, aiming to quickly fill the need by training professionals from other backgrounds to acquire the appropriate skills. A common complaint by professionals attending these programs is the lack of appropriate preparation (in computer science, math or statistics, depending on the student’s background); this is to be expected because typical degrees provide graduate students with expertise narrowly focused to specific areas. Data Science does not fit this paradigm as it is rather interdisciplinary in nature (at least with respect to traditional disciplinary boundaries).

We thus believe that a Data Science undergraduate degree program will be instrumental in educating the future Data Scientists by building their expertise from solid core knowledge, covering the essentials from Computer Science (in managing data) and Statistics (in analysing data) as well as the applications of Data Science in real life problems. A recent search for Data Science undergraduate degrees indicates that there are very few such programs, typically offered as (i) minors to existing B.S. programs (Computer Science, Math or Statistics), (ii) certificate programs, or (iii) programs with very specific focuses (Business Data Analytics or Computational Analytics being the most common ones) [6].

Moreover, because of its ubiquitousness and inherent interdisciplinarity, Data Science has an enormous, and still largely untapped, potential for increasing diversity in computing. A large number of scientific domains synergistic to data science historically have gender-balanced enrollment: indicatively, the departments of

\*All authors contributed equally to this paper.



This work is licensed under a Creative Commons Attribution International 4.0 License.

SIGCSE '21, March 13–20, 2021, Virtual Event, USA.

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8062-1/21/03.

<https://doi.org/10.1145/3408877.3432454>

Astronomy, Biology, and Economics at our institution have consistently observed relatively gender-balanced undergraduate enrollments for the past five years. In contrast, Computer Science enrollment includes 16% female students across its B.S. and B.A. programs [1].

Our main premise in this proposal is that fusing Data Science with core applications from the aforementioned scientific domains can attract student populations from the respective domains to computing, thus enriching diversity. Due to high demand, the data science program will be an appealing alternative for female college applicants who would normally consider careers in natural sciences or even humanities. The recent surge of workshops and conferences that promote diversity in Data Science and related fields, with prominent examples including ‘WiML’ (Women in Machine Learning) [17], ‘WiDS’ (Women in Data Science) [15], and ‘BPDM’ (Broadening Participation in Data Mining) [5] is strong empirical evidence for the validity of our premise.

Moreover, we have observed success at other institutions in creating interdisciplinary programs (including a minor). For example, San Francisco State University created the Promoting Inclusivity in Computing (PINC) program [9], which allows students to complete a minor in ‘Computing Applications.’ The program is targeted toward Biology and Chemistry majors, but all non-CS majors are welcomed to participate. Based on data reported in [9], 60 students have participated in this program, of which 73% are women and 51% are underrepresented minorities (URM).

We draw inspiration from such programs and to this end propose to create a new interdisciplinary undergraduate program in Data Science, offered jointly between the Computer Science and Statistics department.

## 2 MAJOR IN DATA SCIENCE

### 2.1 Motivation

The Computer Science and Engineering (CSE) department at our institution has seen a rising trend in the popularity of CS courses (specifically the introductory course sequence) in recent years. While the CSE department has noticed an increase in the number of students pursuing Computer Science, we have also seen an increased interest and demand for computing courses from non-CS majors. The introductory course sequence offered at our University (CS010 and CS012) introduce students to programming using C++. These courses have very high enrollments, serving students in the College of Engineering (COE), College of Natural and Agricultural Sciences (CNAS), College Of Humanities, Arts, and Social Sciences (CHASS), School of Public Policy (SPP), and School of Business , as well as non-CS students in the College of Engineering. Many students, specifically those majoring in Biology and Bioengineering, elect to take a CS course to satisfy degree requirements. Meanwhile, many other students elect to take CS courses based on interest and curiosity about the field, or recognition that computing is an important skill for the 21st century.

Figure 1 shows the enrollment distribution by College for these two introductory courses for the academic years 2013/14 to 2017/18 (the numbers provided include enrollments for Fall, Winter, Spring, and Summer quarters). For the 2017/18 academic year, more than 73% of enrollments in the introductory course (CS010) are non-majors with 62% of the students from CNAS, as shown by Figure



Figure 1: Enrollments by College

1a. The second course in the sequence (CS012), has approximately 36% non-majors with 25% of students from CNAS, as shown by Figure 1b. We would expect enrollments in CS012 of non-majors to be even higher if the course impact were lifted. Course surveys indicate that while some students take these courses to satisfy a major requirement, the majority are interested to learn about Computer Science and how to fuse their major experience with technical skills.

When examining the enrollments within CNAS for the introductory course CS010, we note the top majors included Mathematics, Biology, Undeclared (within CNAS), Physics, Statistics and Biochemistry, with 34%, 25%, 11%, 10%, 8%, and 5% enrollments, respectively. We recognize that new a degree program must be created to service students who are eager to take computing courses, and ensure that efforts are dedicated to recruitment, mentorship, and curriculum development to support traditionally underrepresented students in computing and technology courses, specifically women.

The Computer Science and Statistics departments have proposed a new *Data Science undergraduate degree program*. This program will be instrumental in educating future data scientists by building knowledge bottom-up, covering both essential knowledge

from Computer Science (in managing data) and Statistics (in analyzing data), and integrating this knowledge with applications to other domains and to real-life problems.

Through its interdisciplinary nature, a Data Science undergraduate program offers a great opportunity to serve as a pathway for attracting more women to computing. Within Computer Science, female enrollment has steadily ranged from 14-16%, as shown by Figure 3, whereas the University overall has about 54% female enrollment.

Majors offered through the College of Natural and Agricultural Sciences have a better gender balance as shown by Figure 2. For 2018, the Statistics department had around 39% female students enrolled in the major. We hope to recruit students into the new major from these more gender-balanced domains.

The proposed program is different from existing programs in that its students will complete custom designed course sequences in other departments (e.g., economics, business, sociology, earth sciences, biology, bioinformatics, and astronomy) where they will learn how Data Science is applied in these domains. Many study areas related to Data Science, such as astronomy, biology, and economics, historically have gender-balanced enrollment. Blending data science with these core applications will help bring these female populations into a broad domain of computing, thus enriching its diversity.

## 2.2 Curriculum

Under the new major, students would choose from which college they wish to have their B.Sc. degree awarded. Students whose degrees are awarded by the College of Engineering would be advised in and have their records maintained by the Department of Computer Science and Engineering; students whose degrees are awarded by the College of Natural and Agricultural Sciences are advised in and have their records maintained by the CNAS.

The new interdisciplinary Data Science program was developed through a committee consisting of faculty from Computer Science, Statistics, as well as advisors from other departments (Biology, Mathematics, Astronomy, Economics, Public Policy). The program is inspired by two recent reports from (i) the National Academies of Sciences, Engineering and Medicine [11], (ii) the Park City Mathematics Institute, a NSF report endorsed by the Board of Directors of the American Statistical Association [7], as well as (iii) existing Data Science programs.

The proposed program consists of 42 courses (quarter-system). Special effort has been taken so that the program satisfies the college general education (GE) requirements for both the College of Engineering (COE) as well as the College of Natural and Agricultural Sciences (CNAS). It includes 8 units of application-area requirements, which is served by taking a **2-course sequence** in another area (outside Computer Science or Statistics). These are selected courses that manipulate, management and analyze data (examples come from biology, economics, electrical engineering, earth sciences, etc.). The General Education (GE) requirements include five courses in Humanities or Social Science (H/SS), three courses in English (ENGL), two Science courses (SCI), one Physical Science course (PHY-SCI) and one Biological Science course (BIO-SCI).

Table 4 highlights course work in Statistics (green), Computer Science (blue), General Education courses (yellow), and application area courses to full-fill the 2-sequence course breath requirement (purple). The course work also includes a set of upper-division technical electives in Computer Science or Statistics, and one on Technology, Policy and Ethics.

Engineering / Ethics

- ENG 170 : Technology, Policy and Ethics

Mathematics

- MATH 9A : First-Year Calculus
- MATH 9B : First-Year Calculus
- MATH 9C : First-Year Calculus
- MATH 10A : Calculus of Several Variables
- MATH 11 : Intro to Discrete Structures
- MATH 31 : Applied Linear Algebra

Computer Science

- CS 10 : Intro to CS for Science, Math, & Eng I
- CS 12 : Intro to CS for Science, Math, & Eng II
- CS 14 : Intro to Data Structures and Algorithms
- CS 100 : Software Construction
- CS 141 : Intermediate Data Structures and Algorithms
- CS 171 : Intro to Machine Learning and Data Mining

Statistics

- STAT 100A : Intro to Statistics
- STAT 100B : Intro to Statistics
- STAT 147 : Intro to Statistical Computing
- STAT 167 : Intro to Data Science
- STAT 170A : Regression Analysis
- STAT 170B : Design of Experiments
- STAT 171 : General Statistical Models

## 2.3 Breath or Application Area Requirements

As mentioned previously, the Major requires a 2-sequence course breath or application area requirement. An example two-course sequence from the Economics Department is *Introductory Econometrics* followed by *Empirical Financial Economics*. Another example using courses from the Business School is *Business Analytics* followed by *Simulation for Business*. Similarly, using Sociology, another sequence is *Population Measures* and *Social Networks*.

These examples are initial sequences for which the prerequisites allow students from the Data Science program to enroll without further coursework. We hope to expand course offerings to other areas and develop new courses that leverage “data munching” and analysis.

## 2.4 Catalog Course Descriptions

The program consists of existing courses from Computer Science and Statistics, breath requirement courses from other application areas, and **four new courses** which are described below.

**2.4.1 CS 105 : Data Analysis Tools (4 units):** Introduction to data analysis tools including data statistics, simple data storage types, data acquisition from the web and public APIs, data cleaning, crowdsourcing for data collection and cleaning, supervised and unsupervised learning techniques, and data visualization. The laboratory will also include hands-on exercises on the aforementioned topics in Python.

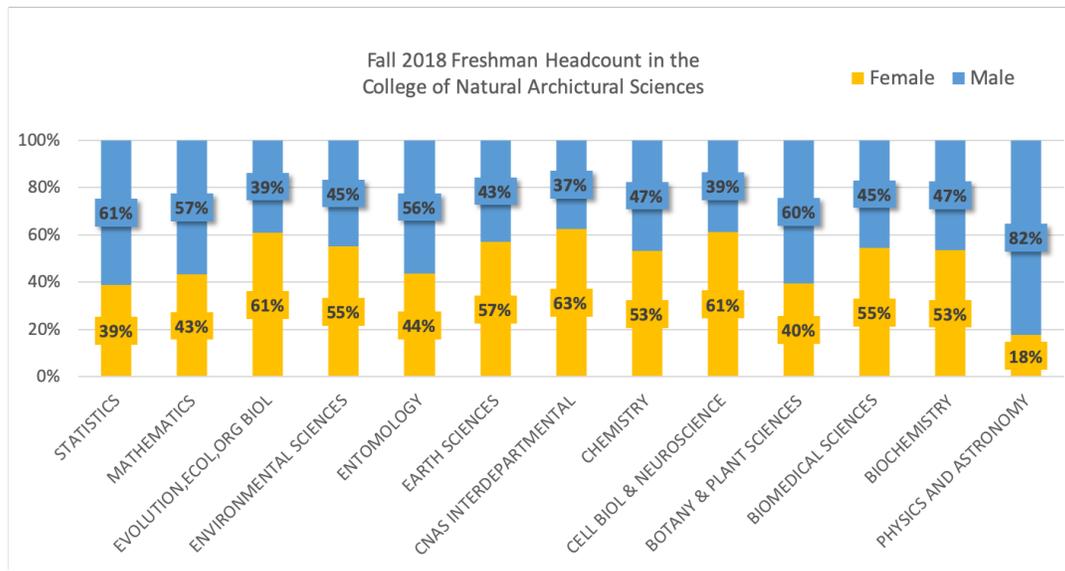


Figure 2: Fall 2018 Fall Headcount in the College of Natural and Agricultural Sciences

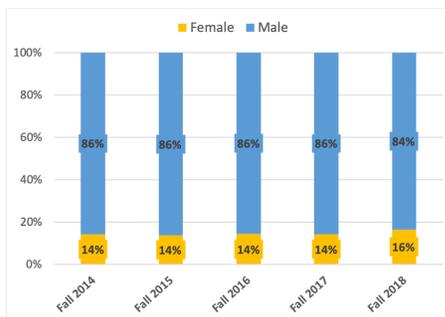


Figure 3: Gender Enrollments in Computer Science (B.S. Degree)

2.4.2 **CS 167: Introduction to Big Data Management (4 units):**

Introduces the design of big-data systems and their application in data management and processing. Describes the common functionality in big-data processing such as distributed storage, resource management, query processing, fault-tolerance, and APIs. Covers the popular big-data technologies such as MapReduce, key-value stores, and semi-structured data management.

2.4.3 **STAT 156A: Mathematical Statistics with Applications for Data Science I (4 units):**

Introduction to frequentist probability concepts, random variables, and their distributions. Discussion of key theorems and inequalities in probability theory. Introduction to frequentist methods of point and interval estimation.

2.4.4 **STAT 156B: Mathematical Statistics with Applications for Data Science II (4 units):**

Introduction to Bayesian probability concepts, illustrative application of Frequentist theory to linear regression, logistic regression and ANOVA, analysis of contingency

Year 1	Fall	Winter	Spring
	MATH 9A	MATH 9B	MATH 9C
	CS 010	CS 012	CS 014
	GE (ENGL)	GE (ENGL)	
	GE (H/SS)	GE (H/SS)	GE (Phy-Sci)
Year 2	Fall	Winter	Spring
	MATH 31	MATH 10A	CS 111
	CS 100	MATH 11	CS 105 (new)
	STAT 100A	STAT 100B	
	GE (Bio-Sci)	GE (Sci)	GE (Sci 2)
Year 3	Fall	Winter	Spring
	STAT 156A (new)	STAT 156B (new)	STAT 167 / CS 171
	CS 167 (new)	CS 141	
	STAT 147	ENGR 170	[TECH ELECTIVE]
	GE (H/SS)	GE (H/SS)	GE (H/SS)
Year 4	Fall	Winter	Spring
	STAT 170A	STAT 170B	Capstone Course
	GE (ENGL)		
	[APP 1]	[APP 2]	
	[TECH ELECTIVE]	[TECH ELECTIVE]	[TECH ELECTIVE]

Figure 4: Data Science Major 4-Year Course Plan

tables, applications of sequential statistics, methods for observational studies and for missing data analyses.

2.4.5 **Justification for New Courses.** When we examined the courses (including technical-electives) offered within Computer Science and Statistics, we found that most topics and skills that are critical for a DS major were already covered by the courses offered. However, we did introduce four new courses to complement the current course work and provide a more in depth dive into relevant topics.

The course titled *Data Analysis Tools* provides a gentle introduction to data analysis and covers tools and methodologies used for the data life-cycle. Students learn to use Python with JupyterHub and learn about the various libraries for data wrangling, cleaning, and analysis (e.g. Matplotlib, Pandas, Sklearn, Numpy, SciPy). This course complements an upper division course titled *Introduction to Machine Learning and Data Mining* and provides the basis for important topics like data exploration and feature engineering.

The second CS course introduced, *Introduction to Big Data Management*, dives into the design of new big data platforms, like Hadoop and Spark. The course will cover systems like Hadoop's Distributed File System and discuss the motivation for the design of such a system and how it achieves scalability and fault-tolerance. While it is important to cover the design of such systems, the course will also ensure students have hand-on experience working with such systems using cloud-platforms, like Amazon's AWS.

We also introduced two new courses in Statistics (Mathematical Statistics with Applications for Data Science I / II). The two courses are fundamental course work in mathematical statistics and offer intuition and modern use of statistical computing tools to illustrate theory by example. The courses focus on specific contexts, rather than abstract general contexts, to show the implications of important theoretical results in mathematical statistics. Upon completing the first course, students will have (i) developed a strong intuition for important theorems, inequalities, and frequentist inference methods in mathematical statistics; (ii) have both a working vocabulary and hands-on experience with both traditional and modern approaches to drawing inferences from data sets; (iii) have experience connecting real-world objectives in data analysis to formal mathematical tools. Upon completing the course, students will have (i) an understanding of the Bayesian paradigm of statistical inference and how it differs from the frequentist paradigm; (ii) seen how statistical theory paves the way for familiar workhorse data analysis methods; (iii) acquired critical data analysis skills for contingency tables, observational studies and missing data scenarios.

**2.4.6 Initial Course Offering.** While the Data Science program is not officially launched until Fall 2020, we decided to pilot two courses in order to collect baseline data and ensure that courses are well-received. We piloted CS 105 : *Data Analysis Tools* in Winter 2020 and Summer 2020, and CS 167 : *Introduction to Big Data Management* in Winter 2020. Both courses employed Project Based Learning (PBL), which required students to work in teams on a large project that included several phases of development.

In CS 105 (Data Analysis Tools), students were asked to select a problem / topic that is relevant to them and formulate several questions to be answered. Once the proposal is submitted and approved, students proceed to work in teams of 2-3 students to (1) collect datasets either via web-scraping or use of APIs, (2) perform Explanatory Data Analysis (EDA), (3) Data Cleaning and Data Integration, and finally (4) Data Analysis. The culmination of this project is a presentation in which students present their Jupyter notebooks (which must include markup and analysis of the presented code / visualizations).

Similarly in CS 167 (Introduction to Big Data Management), students were asked to form a team and select a problem / topic of their choice. The focus of the CS 167 project was on Big Data

processing, hence students had to collect more than 10GB of data and then use Big Data platforms like Hadoop or Spark for data cleaning and analysis. The project also involved several phases that included (1) data collection (2) data filtering and cleaning, (3) data analysis, (4) data visualization / dashboards.

One key observation received from the pilot offering is the importance of clear project rubrics, and the difficulty in creating such rubrics given the variability between projects. It was critical for both course instructors to allow students to select a project topic of their choosing along with the corresponding datasets or data-sources. However, this proved to be a challenge because some projects required more processing in one phase compared to another team's project. Resources like pblworks.org and [12, 16] proved invaluable for creating fair project rubrics and evaluating teams.

The two pilot course enrollment consisted of Computer Science majors, hence the enrollment does not accurately reflect expected enrollment in the future. However, the pilot offering allowed instructors to evaluate the course material, develop project requirements and rubric, and develop a robust course infrastructure. All services for the courses were housed and maintained at the institution to allow flexibility and real-time maintenance. However, we note that as course size increases, platforms like Amazon AWS will be useful to allow increase in course size. Plus, since students will be working with large datasets, they can publicly house these datasets on S3 for free.

## 2.5 Capstone Course

Another defining feature of our major is the Capstone Course, a quarter-long course that involves a team project. Students submissions include a proposal, mid-quarter report, and end-quarter report. In addition, students provide weekly status reports as well as regular code-reviews to ensure they meet the project milestones and deliverables.

The course consists of a one-hour lecture and a three-hour laboratory, per week. The course lecture touches upon case studies in Data Science applications [10] and dives into important topics related to Ethics and data anonymization, and technical reading and writing.

The goal of the project is to provide students with an authentic experience the life-cycle of a Data Science project, which includes **data extraction, data cleaning/manipulation, data exploration, feature engineering, and data analysis**. Projects will be solicited from local companies, research-labs, non-profits and other organizations (e.g. Los Angeles County has a large public data repository). Our motivation is to provide students with a hands-on experience with challenges tasks that a Data Scientist would face day-to-day. Moreover, we ask all our partnering organizations to meet (either in-person or via tele-conference) with students bi-weekly to provide input and direction for the project design. At the end of the course, industry sponsors will be invited to a showcase.

## 3 ADMINISTRATION PLAN

The program will be administered through a joint steering committee. The steering committee will consist of three faculty from the Department of Computer Science and Engineering and three faculty from the Department of Statistics. The program director and the program co-director will be from different departments

and will point-of-contact of the program within their respective departments and colleges. Proposed changes to the program will need to be approved by the majority of the steering committee. The two departments will also cooperate in providing materials needed for any appropriate accreditation process (e.g., ABET [2] or WASC [14].)

As with other intercollege majors, when students declare the new major, they choose from which college they wish to have their B.Sc. degree awarded. Students whose degrees are awarded by the College of Engineering are advised in and have their records maintained by the Department of Computer Science and Engineering; students whose degrees are awarded by the College of Natural and Agricultural Sciences (CNAS) are advised in and have their records maintained by the CNAS Academic Advising Center. Students must fulfill the breadth requirements of the college they choose.

#### 4 ASSESSMENT PLAN

Program assessment and evaluation will consist of reviewing the program objectives and student learning outcomes. Both qualitative and quantitative data will be collected to review the program. The Computer Science and Engineering department, as well as the Statistics department, has well-developed procedures as part of ABET accreditation, hence these best practices will be employed to evaluate the new major. Assessment will be performed at two levels: the course level as well as the program level.

For each CS and STAT course developed or offered as part of the new Major, we will conduct surveys to measure the student's attainment of learning outcomes as well as self efficacy. We acknowledge that traditionally underrepresented students in computing may struggle with imposter syndrome [13]. Pre- and Post- as well as mid-course surveys will be used to evaluate growth mindset of students during the duration of the course and employ appropriate interventions [8].

At the program level, the program Director will evaluate the program by gathering data from all stakeholders, including students, faculty, administrators, and industry partners. Collected data will track student progress through the program and their success in obtaining internships, employment or pursuing graduate school. For example, we will track the number of students (especially women), who graduate from the new program as well as the number of women and minorities who are recruited to the program.

We will also use data from Institutional Research, feedback from industry advisors, student online surveys, and interviews with the Data Science students to analyze the effect of the program. We will iteratively revise our program (teaching pedagogy, structure of capstone project, mentoring plan) based on annual assessments. These program refinements will advance knowledge on how to provide a high-quality interdisciplinary major.

#### 5 RECRUITMENT/OUTREACH

Through our on-going outreach activities, we aim to increase and diversify the recruitment and enrollment of students to our proposed undergraduate program, who may, otherwise, not have been exposed to data science or computing.

Planned outreach activities include (i) teacher training summer institute to provide high-school teachers with content and research exposure in data science, and (ii) a summer bridge program for high

	Applied	Accepted	Currently Enrolled
Female	43 (35%)	38 (37%)	4 (22%)
Male	78 (63%)	63 (61%)	19 (78%)
Other	2 (2%)	2 (2%)	0

school student to build enthusiasm and skills that will encourage students to participate.

The Computer Science department currently has a summer code camp for high-school students that is targeted to increasing the fraction of women and other underrepresented minorities in computing. The code camp introduces students to programming using MIT App Inventor. We hope to expand this program to offer additional camps in Data Science.

#### 6 INITIAL PROGRAM DATA

The program approval process has taken over two years. The lengthy process might be due to its interdisciplinary nature which requires approval from two Departments / Colleges as well as the Academic Senate. Due to delay to getting the program approved, the Data Science program was not 'officially available' for all student in the 2020-2021 academic year. However, students who were accepted into Computer Science, Computer Engineering, Electrical Engineering or Statistics were allowed to switch into this new major. Table 1 summarizes initial program enrollment data.

We recognize the program is quite new and hence enrollment is low (19 students as of July 2020), furthermore, the data is heavily impacted by the COVID-19 situation and skewed since only certain majors were allowed to switch into this new program. Nevertheless, we are hopeful since initial data about applicants include a large number of female students. We hope that these trends will translate to enrollments once the program is officially included in the catalog. We also recognize that outreach and advertisement of this program will play an important role in recruiting a more diverse population of students.

#### 7 CONCLUSION

We have designed an interdisciplinary Data Science major, that is a partnership between the Computer Science and Statistics departments, as a first step toward addressing the deficit of skilled Data Scientist and as a gateway to attract women and underrepresented minorities to computing. We have carefully developed a program grounded in computer science and statistics, with the flexibility to focus on a specific application area (ex. economics, biology, social sciences, etc.). We feel that this major will provide students with the needed skills and opportunities to ultimately secure employment or attend graduate school.

Finally, we believe that the proposed program model and curriculum will serve as a framework for similar academic programs in joint collaboration with statistics, computer science, and faculty in other fields.

#### 8 ACKNOWLEDGEMENTS

We would like to acknowledge the Center for Advancing Women in Technology (CAWIT)'s Technology Pathways Initiative (TPI) for supporting the creation of this program.

## REFERENCES

- [1] [n.d.]. UCR's Institutional Research. [ir.ucr.edu](http://ir.ucr.edu)
- [2] ABET 2019. Criteria for Accrediting Computing Programs (2018-2019). <https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-computing-programs-2018-2019/>. [Online; accessed 19-March-2019].
- [3] Big Data Research and Development (R&D) Initiative [n.d.]. Big Data Research and Development (R&D) Initiative. <https://obamawhitehouse.archives.gov/blog/2016/05/23/administration-issues-strategic-plan-big-data-research-and-development>.
- [4] Big Data to Knowledge (BD2K) Program 2017. Big Data to Knowledge (BD2K) Program. <https://commonfund.nih.gov/bd2k>.
- [5] BPDM 2019. Broadening Participation in Data Mining. <http://www.dataminingshop.com/web/>. [Online; accessed 19-March-2019].
- [6] COLLEGE & UNIVERSITY DATA SCIENCE DEGREES 2017. COLLEGE & UNIVERSITY DATA SCIENCE DEGREES. (<http://datascience.community/colleges>). [Online; accessed 19-March-2019].
- [7] Richard D. De Veaux, Mahesh Agarwal, Maia Averett, Benjamin S. Baumer, Andrew Bray, Thomas C. Bressoud, Lance Bryant, Lei Z. Cheng, Amanda Francis, Robert Gould, Albert Y. Kim, Matt Kretchmar, Qin Lu, Ann Moskol, Deborah Nolan, Roberto Pelayo, Sean Raleigh, Ricky J. Sethi, Mutiara Sondjaja, Neelesh Tiruvilumala, Paul X. Uhlig, Talitha M. Washington, Curtis L. Wesley, David White, and Ping Ye. 2017. Curriculum Guidelines for Undergraduate Programs in Data Science. *Annual Review of Statistics and Its Application* 4, 1 (2017), 15–30. <https://doi.org/10.1146/annurev-statistics-060116-053930> arXiv:<https://doi.org/10.1146/annurev-statistics-060116-053930>
- [8] Carol Dweck. 2015. Carol Dweck Revisits the 'Growth Mindset'.
- [9] Anagha Kulkarni, Ilmi Yoon, Pleuni S. Pennings, Kazunori Okada, and Carmen Domingo. 2018. Promoting Diversity in Computing. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (Larnaca, Cyprus) (ITiCSE 2018)*. Association for Computing Machinery, New York, NY, USA, 236â€241. <https://doi.org/10.1145/3197091.3197145>
- [10] Bernard Marr. 2016. *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. Wiley, Chichester. <https://doi.org/10.1002/9781119278825>
- [11] Engineering National Academies of Sciences and Medicine. 2018. *Data Science for Undergraduates: Opportunities and Options*. The National Academies Press, Washington, DC. <https://doi.org/10.17226/25104>
- [12] Maria Vasilevskaya, David Broman, and Kristian Sandahl. 2015. Assessing Large-Project Courses: Model, Activities, and Lessons Learned. *ACM Trans. Comput. Educ.* 15, 4, Article 20 (Dec. 2015), 30 pages. <https://doi.org/10.1145/2732156>
- [13] Gregory M. Walton and Geoffrey L. Cohen. 2011. A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students. *Science* 331, 6023 (2011), 1447–1451. <https://doi.org/10.1126/science.1198364> arXiv:<https://science.sciencemag.org/content/331/6023/1447.full.pdf>
- [14] WASC 2019. Accrediting Commission for Schools Western Association of Schools and Colleges. <https://www.acswasc.org/>. [Online; accessed 19-March-2019].
- [15] WIDS 2019. Women in Data Science. <https://www.widsconference.org/>. [Online; accessed 19-March-2019].
- [16] Dawn E. Wilkins and Pamela B. Lawhead. 2000. Evaluating Individuals in Team Projects. *SIGCSE Bull.* 32, 1 (March 2000), 172â€175. <https://doi.org/10.1145/331795.331849>
- [17] WIML 2019. Women in Machine Learning. <https://wimlworkshop.org/>. [Online; accessed 19-March-2019].